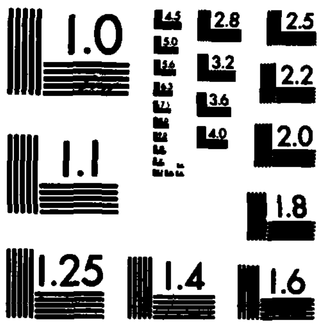
AD-A145 614 ESTIMATION IN NONLINEAR TIME SERIES MODELS I: STATIONARY SERIES..(U) NORTH CAROLINA UNIV AT CHAPEL HILL CENTER FOR STOCHASTIC PREC.. D TJOESTHEIM JUL 80 1/1

UNCLASSIFIED TR-70 AFOSR-TR-84-0826 F49620-82-C-0009 F/G 12/1 NL

END
FILMED



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AFOSR-TR- [illegible] 826

(4)

AD-A145 614

CENTER FOR STOCHASTIC PROCESSES

Department of Statistics
University of North Carolina
Chapel Hill, North Carolina

ESTIMATION IN NONLINEAR TIME SERIES MODELS I: STATIONARY SERIES

by

Dag Tjøstheim

TECHNICAL REPORT #70

July 1980

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) TR #70	5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR- 84-0
6a. NAME OF PERFORMING ORGANIZATION University of North Carolina / 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Air Force Office of Scientific Research
6c. ADDRESS (City, State, and ZIP Code) Department of Statistics Chapel Hill NC 27514	7b. ADDRESS (City, State, and ZIP Code) Directorate of Mathematical & Information Sciences, AFOSR, Bolling AFB DC 20332
8a. NAME OF FUNDING / SPONSORING ORGANIZATION AFOSR / 8b. OFFICE SYMBOL (If applicable) NM	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER F49620-82-C-0009
8c. ADDRESS (City, State, and ZIP Code) Bolling AFB DC 20332	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 61102F; PROJECT NO. 2304; TASK NO. A5; WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)
ESTIMATION IN NONLINEAR TIME SERIES MODELS I: STATIONARY SERIES

12. PERSONAL AUTHOR(S)
Dag Tjostheim

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Technical	FROM ______ TO ______	July 1980	36

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

A general framework for analyzing estimtes in nonlinear time series models is developed. Ergodic strictly stationary series are treated. General conditions for strong consistency and asymptotic normality are derived both for conditional least squares and maximum likelihood type estimates. Examples are taken from exponential autoregressive, random coefficient autoregressive and bilinear time series models. Some nonstationary models and examples are treated in a sequel to this paper.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL CPT Brian W. Woodruff	22b. TELEPHONE (Include Area Code) (202) 767- 5027 / 22c. OFFICE SYMBOL NM

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

ESTIMATION IN NONLINEAR TIME SERIES MODELS I:
STATIONARY SERIES

by

Dag Tjøstheim

Department of Mathematics
University of Bergen
5000 Bergen, Norway

and

Department of Statistics
University of North Carolina
Chapel Hill, North Carolina 27514

Abstract

A general framework for analyzing estimates in nonlinear time series models is developed. Ergodic strictly stationary series are treated. General conditions for strong consistency and asymptotic normality are derived both for conditional least squares and maximum likelihood type estimates. Examples are taken from exponential autoregressive, random coefficient autoregressive and bilinear time series models. Some nonstationary models and examples are treated in a sequel to this paper.

Research supported by AFOSR Contract No. F49620 82 C 0009.

Introduction

Recently there has been a growing interest in nonlinear time series models. Some representative references are Andel (1976) and Nicholls and Quinn (1982) on random coefficient autoregressive models, Granger and Andersen (1978) and Subba Rao (1981) on bilinear models, Haggan and Ozaki (1981) on exponential autoregressive models, Tong and Lim (1980) on threshold autoregressive models, Harrison and Stevens (1976), Ledolter (1981) on dynamic state space models and Priestley (1980) on general state dependent models. A review has been given in Tjøstheim (1984a).

To be able to use nonlinear time series models in practice one must be able to fit the models to data and estimate the parameters. Computational procedures for determining parameters for various model classes are outlined in the above references. Often these are based on a minimization of a least squares or a maximum likelihood type criterion. However, very little is known about the theoretical properties of these procedures and the resulting estimates. An exception is the class of random coefficient autoregressive processes for which a fairly extensive theory of estimation exists (Nicholls and Quinn 1982). See also the special models treated by Robinson (1977) and Aase (1983). Sometimes properties like consistency and asymptotic normality appear to be taken for granted also for other model classes, but some of the simulations performed indicate that there are reasons for being cautious.

In this paper we will try to develop a more systematic approach and discuss a general framework for nonlinear time series estimation. This enables us to survey known results with new proofs as well as to obtain a number of new results. The approach is based on Taylor expansion of a general penalty function which is subsequently

specialized to a conditional least squares and a maximum likelihood type criterion. Klimko and Nelson (1978) have previously considered such Taylor expansions in the conditional least squares case in a general (non-time series) context.

Our approach yields the estimation results of Quinn and Nicholls (1982) as special cases, and, in fact, we are able to weaken their conditions in the maximum likelihood case. The results derived are also applicable to other classes of nonlinear time series. Although the conditions for consistency and asymptotic normality are not always easy to verify, they seem to give a good indication of the specific problems that arise for each class of series. They also suggest that for some models quite strong assumptions could be needed, and thus that there are situations where taking consistency and asymptotic normality for granted may lead astray.

We have found it convenient to subdivide our results into two papers. In the present paper we study strictly stationary ergodic series. This allows us to use the ergodic theorem and the central limit theorem for ergodic strictly stationary martingale differences (Billingsley 1961). The assumption of strict stationarity may appear overly restrictive from a practical point of view, and in some cases it certainly is. However, it should be realized that a strictly stationary nonlinear model is capable of producing realizations with a distinctive nonstationary outlook (cf. e.g. Nicholls and Quinn 1982, Sec. 1 and Tjøstheim 1984a, Sec. 5.1).

An outline of the paper is as follows: In Section 2 we present some results on consistency and asymptotic normality using a general penalty function. In Sections 3 and 5 we specialize to conditional least squares and to a maximum likelihood type penalty function. Applications of our results to a wide range of examples of nonlinear

time series are given in Sections 4 and 6.

In the sequel to this paper (Tjøstheim 1984b) we look at some nonstationary models again basing our results on a Taylor expansion of a general penalty function. A number of additional examples are given in that paper.

2. Two results on consistency and asymptotic normality.

The two results to be stated in this section will be formulated without requiring stationarity, since versions of them will be used also in Tjøstheim (1984b).

Let $\{X_t, t\in I\}$ be a discrete time stochastic process taking values in R^d and defined on a probability space (Ω, F, P). The index set I is either the set Z of all integers or the set N of all positive integers. We assume that observations $(X_1,\ldots,X_n)$ are available. We will treat the asymptotic theory of two types of estimates, namely conditional least squares and maximum likelihood type estimates. Both of these are obtained by minimizing a penalty function, and since, in our setting, the theory is quite similar for the two, we will formulate our results in terms of a general real-valued penalty function $Q_n = Q_n(\beta) = Q_n(X_1,\ldots,X_n;\beta)$ depending on the observations and on a parameter vector β.

The parameter vector $\beta = [\beta_1,\ldots,\beta_r]^T$ will be assumed to be lying in some open set B of Euclidean r-space. Its true value will be denoted by β^0. We will assume that the penalty function Q_n is almost surely twice continuously differentiable in a neighborhood S of β^0. We will denote by $|\cdot|$ the Euclidean norm, so that $|\beta| = (\beta^T\beta)^{\frac{1}{2}}$. For $\delta > 0$, we define $N_\delta = \{\beta: |\beta-\beta^0|< \delta\}$. We will use a.s. as an abbreviation for almost surely, although, when no misunderstanding can arise, it will be omitted in identities involving conditional expectations.

Theorems 2.1 and 2.2 are proved using the standard technique of

Taylor expansion around β^0 (cf. Klimko and Nelson 1978 and Hall and Heyde 1980, Ch. 6). Let $N_\delta \subset S$. Moreover, let $\partial Q_n/\partial\beta$ be the column vector defined by $\partial Q_n/\partial\beta_i$, $i=1,\ldots,r$, and likewise let $\partial^2 Q_n/\partial\beta^2$ be the $r\times r$ matrix defined by $\partial^2 Q_n/\partial\beta_i\partial\beta_j$, $i,j=1,\ldots,r$. Then

$$(2.1)\quad \begin{aligned} Q_n(\beta) &= Q_n(\beta^0) + (\beta-\beta^0)^T \frac{\partial Q_n}{\partial\beta}(\beta^0) + \tfrac{1}{2}(\beta-\beta^0)^T \frac{\partial^2 Q_n}{\partial\beta^2}(\beta-\beta^0) \\ &+ \tfrac{1}{2}(\beta-\beta^0)^T\left\{\frac{\partial^2 Q_n}{\partial\beta^2}(\beta^*) - \frac{\partial^2 Q_n}{\partial\beta^2}(\beta^0)\right\}(\beta-\beta^0) \end{aligned}$$

is valid for $|\beta-\beta^0|<\delta$. Here $\beta^* = \beta^*(X_1,\ldots,X_n;\beta)$ is an intermediate point between β and β^0.

Theorem 2.1: Assume that $\{X_t\}$ and Q_n are such that as $n\to\infty$

A1: $n^{-1}\dfrac{\partial Q_n}{\partial\beta_i}(\beta^0) \overset{a.s.}{\to} 0, \qquad i=1,\ldots,r$

A2: The symmetric matrix $\partial^2 Q_n(\beta^0)/\partial\beta^2$ is non-negative definite and

$$\liminf_{n\to\infty} \lambda^n_{\min}(\beta^0) \overset{a.s.}{>} 0$$

where $\lambda^n_{\min}(\beta^0)$ is the smallest eigenvalue of $\partial^2 Q_n(\beta^0)/\partial\beta^2$.

A3: $$\limsup_{n\to\infty}\,\limsup_{\delta\downarrow 0}\,(n\delta)^{-1}\left|\frac{\partial^2 Q_n}{\partial\beta_i\partial\beta_j}(\beta^*) - \frac{\partial^2 Q_n}{\partial\beta_i\partial\beta_j}(\beta^0)\right| \overset{a.s.}{<} \infty$$

for $i,j=1,\ldots,r$.

Then there exists a sequence of estimators $\hat\beta_n = (\hat\beta_{n1},\ldots,\hat\beta_{nr})^T$ such that $\hat\beta_n \overset{a.s.}{\to} \beta^0$ as $n\to\infty$, and such that for $\varepsilon>0$, there is an event in (Ω,F,P) with $P(E) > 1-\varepsilon$ and an n_0 such that on E and for $n > n_0$, $\partial(Q_n(\hat\beta_n)/\partial\beta_i=0$, $i=1,\ldots,r$, and Q_n attains a relative minimum at $\hat\beta_n$.

Proof: The proof is as in Klimko and Nelson (1978), but it will be

outlined to demonstrate explicitly that the argument does not depend on the special conditional least squares function used there.

Taking into account A1-A3 we can use Egorov's theorem. Thus for a given $\varepsilon > 0$ we can find an $E \epsilon \mathcal{F}$ with $P(E) > 1-\varepsilon$, a positive $\delta^* < \delta$, an $M > 0$, a $\lambda > 0$, and an n_0 such that on E and for $n > n_0$, we have for $\beta \in N_{\delta^*}$

$$|(\beta-\beta^0)^T \frac{\partial Q_n}{\partial \beta}(\beta^0)| < n(\delta^*)^3, \quad |(\beta-\beta^0)^T \frac{\partial^2 Q_n}{\partial \beta^2}(\beta^0)(\beta-\beta^0)| \geq \lambda|\beta-\beta^0|^2 \qquad (2.1)$$

and

$$|(\beta-\beta^0)^T\left\{\frac{\partial^2 Q_n}{\partial \beta^2}(\beta^*) - \frac{\partial^2 Q_n}{\partial \beta^2}(\beta^0)\right\}(\beta-\beta^0)| < nM(\delta^*)^3 \qquad (2.2)$$

Using (2.1) and (2.2) we have that if β is on the boundary of N_{δ^*}, then

$$Q_n(\beta) \geq Q_n(\beta^0) + n(\delta^*)^2(\lambda-\delta^*-M\delta^*), \qquad (2.3)$$

where the last term in (2.3) can be made positive by initially choosing δ sufficiently small. Hence, for such a δ, $Q_n(\beta)$ must attain a minimum at some $\hat{\beta}_n$ in N_{δ^*}, and for this $\hat{\beta}_n$ we must have $\partial Q_n(\hat{\beta}_n)/\partial\beta = 0$. The proof can now be completed as in the proof of Corollary 2.1 of Klimko and Nelson (1978) by selecting appropriate sequences $\{\varepsilon_k\}$ and $\{\delta_k\}$ tending to zero.||

A penalty function Q_n satisfying the general conditions A1-A3 will not necessarily be useful in practice. It seems that additional constraints have to be imposed on the functional form of Q_n to make it natural to choose as $\hat{\beta}_n$ the value of β giving the smallest relative minimum of Q_n. Such properties are inherent in the conditional least squares and maximum likelihood type penalty function. (For e.g. the conditional least squares case we have $E\{Q_n(\beta^0)\} \leq E\{Q_n(\beta)\}$ for all β).

The condition A3 may not always be easy to check in practice. If Q_n is almost surely three times continuously differentiable on S, then, using the mean value theorem, an obvious sufficient condition for A3 is the existence of an $M>0$ independent of β such that

$$\text{A3}': \quad \limsup_{n\to\infty} n^{-1} \left| \frac{\partial^3 Q_n}{\partial\beta_i \partial\beta_j \partial\beta_k} \right| \overset{\text{a.s.}}{<} M$$

for $\beta \in S$ and $i,j,k = 1,\ldots,r$.

When it comes to asymptotic normality it is essentially sufficient to prove asymptotic normality of $\partial Q_n(\beta^0)/\partial\beta$.

<u>Theorem 2.2</u>: Assume that the conditions of Theorem 2.1 are fulfilled and that in addition we have that as $n \to \infty$

$$\text{B1}: \quad n^{-1} \frac{\partial^2 Q_n}{\partial\beta_i \partial\beta_j}(\beta^0) \overset{\text{a.s.}}{\to} V_{ij}$$

for $i,j = 1,\ldots,r$, where $V = (V_{ij})$ is a strictly positive definite matrix, and

$$\text{B2}: \quad n^{-\frac{1}{2}} \frac{\partial Q_n}{\partial\beta}(\beta^0) \overset{d}{\to} N(0,W)$$

where $N(0,W)$ is used to denote a multivariate normal distribution with a zero mean vector and covariance matrix W. Let $\{\hat{\beta}_n\}$ be the estimators obtained in Theorem 2.1. Then

$$n^{\frac{1}{2}}(\hat{\beta}_n - \beta^0) \overset{d}{\to} N(0, V^{-1}WV^{-1}) \tag{2.4}$$

The proof is identical to the proof of Theorem 2.2 of Klimko and Nelson (1978) and is therefore omitted.

In the remaining part of this paper $\{X_t\}$ will be assumed to be strictly stationary and ergodic. In addition second moments of $\{X_t\}$ will always be assumed to exist, so that $\{X_t\}$ is second order stationary as well. The task of finding nonlinear models satisfying these assumptions is far from trivial (cf. Tjøstheim 1984a, Sec. 5).

3. Conditional Least Squares

We denote by F_t^X the sub σ-field of F generated by $\{X_s,\ s\le t\}$, and we will use the notation $\tilde{X}_{t|t-1} = \tilde{X}_{t|t-1}(\beta)$ for the conditional

expectation $E_\beta(X_t | F^X_{t-1})$. We will often omit β for notational convenience. Since second moments of $\{X_t\}$ are assumed to exist, $\tilde{X}_{t|t-1}$ is the optimal one-step least squares predictor of X_t and $E(X_t - \tilde{X}_{t|t-1})^2$ is finite.

In the case where $\{X_t\}$ is defined for $t \geq 1$ only (this will be referred to as the one sided case), $\tilde{X}_{t|t-1}$ will in general depend explicitly on t and therefore $\tilde{X}_{t|t-1}$ do not define a stationary process. If the index set I of $\{X_t, t \in I\}$ comprises all the integers, then $\tilde{X}_{t|t-1}$ is stationary, but in general $\tilde{X}_{t|t-1}$ will depend on X_t's not included in the set of observations $(X_1, \ldots, X_n)$. To avoid these problems we replace F^X_{t-1} by $F^X_{t-1}(m)$, which is the σ-field generated by $\{X_s, t-m \leq s \leq t-1\}$, and let $\tilde{X}_{t|t-1} = E\{X_t | F^X_{t-1}(m)\}$. Here m is an integer at our disposal, and we must have $t \geq m+1$ in the one sided case.

We will use the penalty function

$$Q_n(\beta) = \sum_{t=m+1}^{n} \{X_t - \tilde{X}_{t|t-1}(\beta)\}^2 \tag{3.1}$$

and the conditional least squares estimates will be obtained by minimizing this function. In the important special case where $\tilde{X}_{t|t-1}$ only depends on $\{X_s, t-p \leq s \leq t-1\}$, i.e. $\{X_t\}$ is a nonlinear autoregressive process of order p, we can take m=p and we have $E(X_t | F^X_{t-1}) = E\{X_t | F^X_{t-1}(m)\}$, where $t \geq m+1$ in the one sided case. It should be noted that we may loose parameters of interest in the conditioning operation with respect to F^X_{t-1} or $F^X_{t-1}(m)$, but as will be shown in the examples of the next section, there are ways of getting around this problem.

The theorems in this section are essentially obtained by reformulating and extending the arguments of Klimko and Nelson (1978) to the multivariate case.

<u>Theorem 3.1</u>: Assume that $\{X_t\}$ is a d-dimensional strictly stationary ergodic process with $E(|X_t|^2 < \infty$ and such that $\tilde{X}_{t|t-1}(\beta) = E_\beta\{X_t | F^X_{t-1}(m)\}$ is almost surely three times continuously differentiable in an open set B containing β^0. Moreover, suppose that

C1: $E\left\{\left|\frac{\partial \tilde{X}_{t|t-1}}{\partial \beta_i}(\beta^o)\right|^2\right\} < \infty$ and $E\left\{\left|\frac{\partial^2 X_{t|t-1}}{\partial \beta_i \partial \beta_j}(\beta^o)\right|^2\right\} < \infty$

for $i,j = 1,\ldots,r$.

C2: The vectors $\partial \tilde{X}_{t|t-1}(\beta^o)/\partial \beta_i$, $i = 1,\ldots,r$, are linearly independent in the sense that if $a_1,\ldots,a_r$ are arbitrary real numbers such that $E\left\{\left|\sum_{i=1}^{r} a_i \frac{\partial \tilde{X}_{t|t-1}}{\partial \beta_i}(\beta^0)\right|^2\right\} = 0$, then $a_1 = a_2 = \ldots = a_r = 0$.

C3: For $\beta \in B$, there exists functions $G^{ijk}_{t-1}(X_1,\ldots,X_{t-1})$ and $H^{ijk}_t(X_1,\ldots,X_t)$ such that

$$\left|\frac{\partial \tilde{X}_{t|t-1}}{\partial \beta_i}(\beta)\frac{\partial^2 \tilde{X}_{t|t-1}}{\partial \beta_j \partial \beta_k}(\beta)\right| \le G^{ijk}_{t-1}, \quad E(G^{ijk}_{t-1}) < \infty$$

$$\left|\{X_t - \tilde{X}_{t|t-1}(\beta)\}\frac{\partial^3 \tilde{X}_{t|t-1}}{\partial \beta_i \partial \beta_j \partial \beta_k}(\beta)\right| \le H^{ijk}_t, \quad E(H^{ijk}_t) < \infty$$

for $i,j,k=1,\ldots,r$.

Then there exists a sequence of estimators $\{\hat{\beta}_n\}$ minimizing Q_n of (3.1) such that the conclusion of Theorem 2.1 holds.

<u>Proof</u>: Using (3.1) we have

$$\frac{\partial Q_n}{\partial \beta_i} = -2\sum_{t=m+1}^{n} (X_t - \tilde{X}_{t|t-1})^T \frac{\partial \tilde{X}_{t|t-1}}{\partial \beta_i} \triangleq -2\sum_{t=m+1}^{n} \psi_t \tag{3.2}$$

But

$$E\{\psi_t(\beta^0)\} = E\left\{E\left[\{X_t - X_{t|t-1}(\beta^0)\}^T | F^X_{t-1}(m)\right] \cdot \frac{\partial \tilde{X}_{t|t-1}}{\partial \beta_i}(\beta^0)\right\} = 0 \tag{3.3}$$

Furthermore, from the Schwarz inequality and C1 we have $E\{|\psi_t(\beta^0)|\} < \infty$ and using the ergodic theorem we have that A1 of Theorem 2.1 is fulfilled.

Second order derivatives are given by

$$\frac{\partial^2 Q_n}{\partial \beta_i \partial \beta_j} = 2\sum_{t=m+1}^{n} \frac{\partial \tilde{X}_{t|t-1}}{\partial \beta_i}^T \frac{\partial \tilde{X}_{t|t-1}}{\partial \beta_j} - 2(X_t - \tilde{X}_{t|t-1})^T \frac{\partial^2 \tilde{X}_{t|t-1}}{\partial \beta_i \partial \beta_j}. \tag{3.4}$$

Reasoning exactly as above we have that the expectation of the last term of (3.4) is zero for $\beta = \beta^0$. Again using the Schwarz inequality and C1 it follows from the ergodic theorem that as $n \to \infty$,

$$n^{-1} \frac{\partial^2 Q_n}{\partial \beta_i \partial \beta_j}(\beta^0) \xrightarrow{a.s.} 2E\left\{\frac{\partial \tilde{X}_{t|t-1}}{\partial \beta_i}(\beta^0) \frac{\partial \tilde{X}_{t|t-1}}{\partial \beta_j}(\beta^0)\right\} \triangleq V_{ij} \tag{3.5}$$

for $i,j=1,\ldots,r$. The matrix $V = (V_{ij})$ in (3.5) is by definition non-negative definite. That its smallest eigenvalue is larger than zero follows from C2, and this in turn implies that A2 of Theorem 2.1 is fulfilled. Finally, from C3 we have that the ergodic theorem and the mean value theorem implies A3 of Theorem 2.1 ||

Let $\partial \tilde{X}_{t|t-1}/\partial \beta$ be the $d \times r$ matrix having $\partial \tilde{X}_{t|t-1}/\partial \beta_i$, $i=1,\ldots r$, as its column vectors. We denote by $U = \frac{1}{2}V$ the $r \times r$ matrix defined by

$$U = E\left\{\frac{\partial \tilde{X}_{t|t-1}}{\partial \beta}^T(\beta^0) \frac{\partial \tilde{X}_{t|t-1}}{\partial \beta}(\beta^0)\right\}. \tag{3.6}$$

We will next use Theorem 2.2 to prove asymptotic normality. The proof depends on Billingsley's (1961) martingale central limit theorem. We then need to condition with respect to an increasing sequence of σ-fields in order to obtain a martingale, and since $\{F^X_t(m)\}$ is not increasing, we now assume the existence of an m such that we have

($t \geq m+1$ in one sided case)

$$E(X_t | F^X_{t-1}) \overset{a.s.}{=} E(X_t | F^X_{t-1}(m))$$

and (3.7)

$$f_{t|t-1} \overset{\Delta}{=} E\{(X_t - X_{t|t-1})(X_t - X_{t|t-1})^T | F^X_{t-1}\}$$

$$\overset{a.s.}{=} E\{(X_t - X_{t|t-1})(X_t - X_{t|t-1})^T | F^X_{t-1}(m)\}$$

where we have used $f_{t|t-1}$ to denote the d×d conditional prediction error matrix of $\{X_t\}$. The relations in (3.7) hold trivially for nonlinear AR processes.

Theorem 3.2: Assume that (3.7) and the conditions of Theorem 3.1 are fulfilled. In addition assume that

D1: $$R = E\left\{\frac{\partial \tilde{X}_{t|t-1}^T}{\partial \beta}(\beta^0)\, f_{t|t-1}(\beta^0)\, \frac{\partial \tilde{X}_{t|t-1}}{\partial \beta}(\beta^0)\right\} < \infty$$

Let $\{\hat{\beta}_n\}$ be the estimators obtained in Theorem 3.1. Then

$$n^{\frac{1}{2}}(\hat{\beta}_n - \beta^0) \xrightarrow{d} N(0, U^{-1}RU^{-1}). \tag{3.8}$$

Proof: In view of Theorem 3.1 and (3.5) we only have to verify condition B2 of Theorem 2.2. This will be done using a Cramer-Wold type argument. Thus let $\gamma_1, \ldots, \gamma_r$ be arbitrary real numbers. Using the definition of $\tilde{X}_{t|t-1}$ we have

$$E\left|\sum_{i=1}^r \gamma_i \frac{\partial \tilde{X}_{t|t-1}^T}{\partial \beta_i}(\beta^0)\{X_t - \tilde{X}_{t|t-1}(\beta^0)\} | F^X_{t-1}\right| = 0 \tag{3.9}$$

and from (3.2) we conclude that the time increments of $\sum_{i=1}^r \gamma_i \partial Q_n(\beta^0/\partial\beta_i)$ are strictly stationary ergodic martingale increments with respect to $\{F^X_t\}$. It follows from Billingsley (1961) that $n^{-\frac{1}{2}}\sum_{i=1}^r \gamma_i \partial Q_n(\beta^0)/\partial\beta_i$ converges in law to a normal distribution with zero mean, and thus $n^{-\frac{1}{2}}\partial Q_n(\beta^0)/\partial\beta_i$ has a multivariate normal distribution as its limiting

distribution.

It remains to evaluate the covariance matrix. Since $\{\partial Q_n(\beta^0)/\partial\beta_i, F_n^X\}$ is a martingale, using (3.2) it is easy to verify that

$$E\left\{\frac{\partial Q_n}{\partial\beta_i}(\beta^0)\,\frac{\partial Q_n}{\partial\beta_j}(\beta^0)\right\} =$$

$$4\sum_{t=1}^{n} E\left(\frac{\partial\tilde{X}^T_{t|t-1}}{\partial\beta_i}(\beta^0)E\left[\{X_t-\tilde{X}_{t|t-1}(\beta^0)\}\{X_t-\tilde{X}_{t|t-1}(\beta^0)\}^T|F^X_{t-1}\right]\frac{\partial\tilde{X}_{t|t-1}}{\partial\beta_j}(\beta^0)\right)$$

and by combining (2.4), (3.5), (3.6), (3.7) and D1 we obtain (3.8).||

For a large class of time series models (including the ordinary linear AR models) the condition D1 is implied by the condition C1 of Theorem 3.1, and hence essentially no extra condition is required to ensure asymptotic normality.

Corollary 3.1 : If $X_t-\tilde{X}_{t|t-1}(\beta^0)$ is independent of F^X_{t-1}, then D1 is implied by C1.

Proof: Under the stated independence assumption we have

$$f_{t|t-1}(\beta^0) = E\left[\{X_t-X_{t|t-1}(\beta^0)\}\{X_t-X_{t|t-1}(\beta^0)\}^T\right] \tag{3.11}$$

and the Schwarz inequality yields the conclusion.||

4. Examples

A number of models was referred to in the Introduction. In Tjøstheim (1984a) these were subdivided into three main categories: Models motivated by nonlinear differential equations, bilinear models and doubly stochastic models. We will try to apply the results obtained in the preceding section to one model from each category, namely an exponential autoregressive model, a bilinear model and a random coefficient autoregressive series.

For notational convenience we will omit the superscript 0 for the

true value of the parameter vector in this section. Moreover, in all of the following $\{e_t, -\infty < t < \infty\}$ will denote a sequence of independent identically distributed (iid) (possibly vector) random variables with $E(e_t) = 0$ and $E(e_t e_t^T) = G < \infty$.

4.1 Exponential autoregressive models

These models were introduced and studied by Ozaki (1980) and Haggan and Ozaki (1981). The point of departure is an ordinary scalar autoregressive model of order p (AR(p)), where the ith autoregressive coefficient a_i is replaced by $a_i(t-1,X) = \psi_i + \pi_i \exp(-\gamma X_{t-1}^2)$, where $\psi_i, \pi_i, i=1,\ldots,p$ and γ are real parameters such that $\gamma > 0$. This results in a model

$$X_t - \sum_{i=1}^{p} \{\psi_i + \pi_i \exp(-\gamma X_{t-1}^2)\} X_{t-i} = e_t \tag{4.1}$$

which is assumed defined for $t \geq p+1$ with $X_1,\ldots,X_p$ being initial variables.

Haggan and Ozaki (1981) have considered the problem of numerical evaluation of $\hat{\psi}_{in}$, $\hat{\pi}_{in}$ and $\hat{\gamma}_n$ by minimization of the sum of squares penalty function Q_n of (3.1), and they have done simulations. However, we are not aware of any results concerning the asymptotic properties of these estimates.

To make the principles involved more transparent we will work with the first order model

$$X_t - \{\psi + \pi \exp(-\gamma X_{t-1}^2)\} X_{t-1} = e_t \tag{4.2}$$

defined for $t \geq 2$ with X_1 being an initial variable.

<u>Theorem 4.1</u>: Let $\{X_t\}$ be defined by (4.2). Assume that $|\phi| + |\pi| < 1$, and that e_t has a density function with infinite support such that $E(e_t^6) < \infty$. Then there exists a unique distribution for the initial

variable X_1 such that $\{X_t,\ t \geq 1\}$ is strictly stationary and ergodic. Moreover, there exists a sequence of estimators $\{[\hat{\psi}_n, \hat{\pi}_n, \hat{\gamma}_n]\}$ minimizing (as described in the conclusion of Theorem 2.1) the penalty function Q_n of (3.1) and such that $[\hat{\psi}_n, \hat{\pi}_n, \hat{\gamma}_n] \xrightarrow{a.s.} [\psi, \pi, \gamma]$, and $[\hat{\psi}_n, \hat{\pi}_n, \hat{\gamma}_n]$ is asymptotically normal.

Proof: Our independence assumption on $\{e_t\}$ implies that $\{X_t,\ t \geq 1\}$ is a Markov process, and the problem of existence of a strictly stationary and ergodic solution to the difference equation (4.2) can then be treated using Corollary 5.2 of Tweedie (1975).

Since e_t has a density with infinite support it follows that $\{X_t\}$ is ϕ-irreducible (cf. Tweedie 1975) with ϕ being Lebesgue measure. Since for an arbitrary Borel set B we have

$$P(x,B) = P\,(X_t \in B \mid X_{t-1} = x) = P\,(e_t \in B - a(x)\cdot x) \tag{4.3}$$

where $a(x) = \psi + \pi \exp(-\gamma x^2)$, and since the function a is continuous, it follows that $\{P(x,\cdot)\}$ is strongly continuous. Moreover, it is easily seen from (4.2) that

$$\gamma_x = E\{(|X_t| - |X_{t-1}|) \mid X_{t-1} = x\} \leq \{|a(x)| - 1\}\ |x| + E|e_t|. \tag{4.4}$$

Here, $|a(x)| \leq |\psi| + |\pi| \exp(-\gamma x^2) \leq |\psi| + |\pi|$ since $\gamma \geq 0$. Let $\alpha = E(|e_t|)/(1 - |\psi| - |\pi|)$. Then if $|\psi| + |\pi| < 1$, there exists a $c > 0$ such that $\gamma_x \leq -c$ for all x with $|x| > \alpha$. Moreover, γ_x is bounded from above for all x with $|x| \leq \alpha$. It follows from Corollary 5.2 of Tweedie (1975) that there exists a unique invariant initial distribution for X_1 such that $\{X_t,\ t \geq 1\}$ is strictly stationary and ergodic.

Since we have a nonlinear AR(1) process, we can take m=1 in Theorems 3.1 and 3.2. The conditions stated in (3.7) will then be trivially fulfilled and we have for $t \geq 2$

$$\tilde{X}_{t|t-1} = E(X_t|F^X_{t-1}) = \{\psi + \pi \exp(-\gamma X^2_{t-1})\}X_{t-1}. \quad (4.5)$$

Furthermore, $f_{t|t-1} = E(e_t^2) = \sigma^2$ such that D1 of Theorem 3.2 follows from C1 of Theorem 3.1, and it is sufficient to verify C1 - C3.

Since any moment of $\pi \exp(-\gamma X^2_{t-1})X_{t-1}$ exists, it follows from (4.2) and the strict stationarity of $\{X_t\}$ that $E(e_t^6) < \infty$ implies $E(X_t^6) < \infty$. From (4.5) we have

$$\frac{\partial \tilde{X}_{t|t-1}}{\partial \psi} = X_{t-1}, \quad \frac{\partial \tilde{X}^k_{t|t-1}}{\partial \gamma^k} = (-2)^k \pi \exp(-\gamma X^2_{t-1})X^{k+1}_{t-1}$$

$$\frac{\partial \tilde{X}_{t|t-1}}{\partial \pi} = \exp(-\gamma X^2_{t-1})X_{t-1}, \frac{\partial \tilde{X}^{k+1}_{t|t-1}}{\partial \gamma^k \partial \pi} = (-2)^k \exp(-\gamma X^2_{t-1})X^{k+1}_{t-1} \quad (4.6)$$

for k=1,... while the other derivatives are zero. It is easily seen that $E(X_t^6) < \infty$ implies that C1 is satisfied. Since $|\psi| + |\pi| < 1$, we have that $|X_t - \tilde{X}_{t|t-1}| \leq |X_t| + |X_{t-1}|$ and that the above derivatives are bounded by $|X_{t-1}|$, $2^k|X_{t-1}|^{k+1}$, $|X_{t-1}|$ and $2^k|X_{t-1}|^{k+1}$, respectively. Successive applications of the Schwarz inequality and use of $E(X_t^6) < \infty$ yield C3.

Let a_1, a_2 and a_3 be three arbitrary real numbers. Then

$$E\left(\left|a_1 \frac{\partial \tilde{X}_{t|t-1}}{\partial \psi} + a_2 \frac{\partial \tilde{X}_{t|t-1}}{\partial \pi} + \frac{\partial \tilde{X}_{t|t-1}}{\partial \gamma}\right|^2\right) = 0 \quad (4.7)$$

implies

$$X_{t-1}\left[a_1 + \exp(-\gamma X^2_{t-1})\{a_2 X_{t-1} - 2a_3\pi\}\right] \overset{a.s.}{=} 0, \quad (4.8)$$

and since $E(X_t^2) \geq E(e_t^2) > 0$, it follows that $a_1 = a_2 = a_3 = 0$. Hence C2 holds and the proof is completed. ||

The infinite support assumption on $\{e_t\}$ can be relaxed. Moreover, it is not absolutely critical that the model (4.2) is initiated with X_1 in its stationary invariant distribution. The

critical fact is the existence of such a distribution (cf. Klimko and Nelson 1978, Sec. 4).

The general P-th order model can be transformed to a first order vector autoregressive model, and essentially the same technique can be used. In this case, the condition $|\psi| + |\pi| < 1$ can be replaced by the condition that there is a matrix norm $|\cdot|$ such that $|\Psi| + |\Pi| < 1$, where

$$\Psi = \begin{vmatrix} \Psi_{p-1} & \psi_p \\ I_{p-1} & 0 \end{vmatrix} \qquad \Pi = \begin{vmatrix} \Pi_{p-1} & \pi_p \\ I_{p-1} & 0 \end{vmatrix}$$

with $\psi_{p-1} = [\psi_1,\ldots,\psi_{p-1}]$ and $\Pi_{p-1} = [\pi_1,\ldots,\pi_{p-1}]$, and where I_{p-1} is the identity matrix of order p-1.

It is interesting to consider the special case of an ordinary AR(p) process. Then $\gamma=0$ and $\psi_i + \pi_i = a_i$. The ergodicity and stationarity condition from Tweedie (1975) reduces to requiring that the zeros of $z^p - \sum_{i=1}^{p} a_i z^{p-i}$ are inside the unit circle of the complex z-plane. Since only first order derivatives of $\tilde{X}_{t|t-1}$ are non-zero, and since $f_{t|t-1} = E(e_t^2) = \sigma^2$ and X_{t-i}, $i=1,\ldots,p$, are linearly independent, we have that C1-C3 and D1 amount to requiring $E(X_t^2) < \infty$, or $E(e_t^2) < \infty$, so that Theorems 3.1 and 3.2 reduce to the classical consistency and central limit theorem in this situation.

A related class of models is the threshold autoregressive processes (Tong and Lim 1980). Unfortunately we have not been able to establish the existence of a stationary invariant initial distribution for these processes. The transition probability $P(x,\cdot)$ is not in general strongly continuous (nor is it weakly continuous), and this makes it

difficult to apply Tweedie's (1975) criterion. We will treat the threshold processes in Tjøstheim (1984b), however.

Another class of related processes is studied by Aase (1984) (see also Jones 1978). In particular Aase looks at models of type

$$X_t - \theta f(X_{t-1}) = e_t \tag{4.10}$$

The parameter θ is to be estimated. Clearly

$$\tilde{X}_{t|t-1} = \theta f(X_{t-1}) \text{ and } \frac{\partial \tilde{X}_{t|t-1}}{\partial\theta} = f(X_{t-1}) \tag{4.11}$$

while higher order derivatives are zero. Using the method of the preceding proof it is not difficult to show that if e_t has a density with an infinite support and $E(e_t^2) < \infty$, if f is continuous, non-zero almost everywhere, and there is a constant c such that $|f(x)| \leq c|x|$, and if $|\theta| < c^{-1}$, then the estimate $\hat{\theta}_n = \sum_{t=2}^{n} X_t f(X_{t-1}) / \sum_{t=2}^{n} f^2(X_{t-1})$ is strongly consistent and asymptotically normal. Some other examples are discussed by Aase who uses densities with bounded support and recursively defined estimates.

4.2 Random coefficient autoregressive (RCA) models.

These are defined by allowing random additive pertubations of the AR coefficients of ordinary AR models. Thus a d-dimensional RCA model of order p is defined by

$$X_t - \sum_{i=1}^{p} (a_i + b_{ti}) X_{t-i} = e_t \tag{4.12}$$

for $-\infty<t<\infty$. Here, a_i, $i=1,\ldots,p$, are deterministic $d \times d$ matrices, whereas $\{b_t(p)\} = \{[b_{t1},\ldots,b_{tp}]\}$ defines a $d \times pd$ zero-mean matrix process with the $b_t(p)$'s being iid and independent of $\{e_t\}$. Second moments of both $\{b_t(p)\}$ and $\{e_t\}$ will be assumed to exist, and the covariance matrix $G = E(e_t e_t^T)$ will be assumed to be nonsingular. We

denote by F_t^e and F_t^b the σ-fields generated by $\{e_s, s \le t\}$ and $\{b_s(p), s \le t\}$, and we assume that conditions are fulfilled so that an ergodic, strictly and second order stationary $F_t^e \vee F_t^b$-measurable solution of (4.12) exists. Various conditions for this in terms of the matrices a_i, $i=1,\ldots,p$, and the second moments of $\{b_t(p)\}$ are given in Nicholls and Quinn (1982, Ch. 2).

A least squares estimation theory of RCA processes is developed in Chs. 3 and 7 of Nicholls and Quinn (1982) using regression analysis and conditioning with respect to $F_t^e \vee F_t^b$. We will demonstrate that the general framework of estimation developed in this paper can be used to obtain their results, i.e. their theorems 3.1, 3.2, 7.1 and 7.2.

It is convenient to rewrite (4.12) slightly. Let $b_t^T = [\text{vec}^T(b_{t1}),\ldots,\text{vec}^T(b_{tp})]$, where $\text{vec}(b_{ti})$ is the d^2-dimensional column vector obtained by stacking the columns of b_{ti} one on top of the other in order from left to right, and let $a^T = \left(\text{vec}^T(a_1),\ldots,\text{vec}^T(a_p)\right)$. Moreover, let c_1, c_2 and c_3 be matrices such that the product $c_1c_2c_3$ is well defined. Applying the formula $\text{vec}(c_1c_2c_3) = (c_3^T \otimes c_1)\text{vec}(c_2)$, where $\otimes$ denotes tensor product, we obtain using vectorization on (4.12)

$$X_t - F(t-1,X)\,(a + b_t) = e_t. \tag{4.13}$$

Here $F(t-1,X)$ is the $d \times pd^2$ matrix function given by

$$F(t-1,X) = [X_{t-1}^T \otimes I_d, \ldots, X_{t-p}^T \otimes I_d] . \tag{4.14}$$

We describe the processes $\{b_t\}$ and $\{e_t\}$ by their covariance matrices Λ and G, respectively. This amounts to a complete description in the Gaussian case. The parameters of interest are then the pd^2 elements of the vector a and the $pd^2(pd^2 + 1)/2 + d(d + 1)/2$ distinct elements

of the symmetric matrices Λ and G.

Theorem 4.2: Under the stated assumptions on $\{X_t\}$ there exists a strongly consistent sequence of estimates $\{\hat{a}_n\}$ for a. If we assume in addition that $E\{X_{ti}^4\} < \infty$, i=1,...,d, where X_{ti} is the ith component of X_t, and that e_t cannot take on only two values almost surely, then there exists strongly consistent sequences of estimates $\{\hat{\Lambda}_n\}$ and $\{\hat{G}_n\}$ for Λ and G.

Proof: Note that (3.7) is fulfilled if we take m = p, so that we may condition on F_{t-1}^X instead of $F_{t-1}^X(m)$ both in this proof and in the proof of asymptotic normality in the next theorem.

Since $\{X_t\}$ is generated by $\{e_t\}$ and $\{b_t\}$ we have $F_t^X \subset F_t^b \vee F_t^e$, and due to our independence and zero mean assumptions

$$E\{F(t-1,X)b_t|F_{t-1}^X\} = F(t-1,X)E\{E(b_t|F_{t-1}^b \vee F_{t-1}^e)|F_{t-1}^X\} = 0. \qquad (4.15)$$

It follows immediately that $\tilde{X}_{t|t-1} = F(t-1,X)a$. The equation $\partial Q_n/\partial vec(a) = 0$ with Q_n as in (3.1) is linear in a, and an explicit and unique solution $\hat{a}_n$ can be found (cf. Quinn and Nicholls 1982, p.126).

Using vector notation in derivatives we have $\partial\tilde{X}_{t|t-1}/\partial vec(a_i) = X^T(t-i) \otimes I_d$ while the higher order derivatives with respect to a_i, i=1,...,p, are zero. Since by assumption $\{X_t\}$ has second moments, it follows at once that C1 and C3 of Theorem 3.1 are fulfilled. The linear independence condition C2, on the other hand, follows from the nonsingularity of the matrix G (cf. Nicholls and Quinn 1982, proof of Th. 2.2, p. 24), and the first part is proved.

To obtain estimates of Λ and G we use the conditional least squares principle with X_t replaced by

$$v_t = \{X_t-\tilde{X}_{t|t-1}(a)\}\{X_t-\tilde{X}_{t|t-1}(a)\}^T = (F(t-1,X)b_t + e_t)\ F(t-1,X)b_t+e_t)^T \quad (4.16)$$

Using the same reasoning as when deriving (4.15) we have

$$f_{t|t-1} = \tilde{v}_{t|t-1} = E(v_t|F^X_{t-1}) = F(t-1,X)\ \Lambda\ F^T(t-1,X) + G. \qquad (4.17)$$

We are only interested in the $pd^2(pd^2 + 1)/2 + d(d + 1)/2$ distinct elements of Λ and G, and we therefore use the vech operation on (4.17). We refer to Nicholls and Quinn(1982, Ch. 1) for a description of this operation. We have

$$\text{vech}(\tilde{v}_{t|t-1}) = H\{F(t-1,X) \otimes F(t-1,X)\}K^T\ \text{vech}(\Lambda)+\text{vech}(G) \qquad (4.18)$$

where H and K are constant matrices independent of Λ and G.

We now consider conditional least squares estimates obtained by minimizing

$$Q_n'(v) = \sum_{t=p+1}^{n} |\text{vech}(v_t) - \text{vech}(\tilde{v}_{t|t-1})|^2 \qquad (4.19)$$

where in practive v_t has to be replaced by

$\hat{v}_t = \{X_t-\tilde{X}_{t|t-1}(\hat{a}_n)\}\{X_t-X_{t|t-1}(\hat{a}_n)\}^T$. From (4.18) it follows that

$$\frac{\partial\ \text{vech}(\tilde{v}_{t|t-1})}{\partial\ \text{vech}\ (\Lambda)} = H\{F(t-1,X) \otimes F(t-1,X)\}K \text{ and } \frac{\partial\ \text{vech}(\tilde{v}_{t|t-1})}{\partial\ \text{vech}\ (G)} = I_{d(d-1)/2} \quad (4.20)$$

while higher order derivatives with respect to the parameters are zero. The fact that e_t cannot take on only two values almost surely now ensures (cf. Nicholls and Quinn 1982, p.45 and their Lemma 3.1) that C2 of Theorem 3.1 holds, while (4.20) and the existence of fourth order moments of $\{X_t\}$ means that C1 and C3 are fulfilled, so that the estimation of $\hat{\Lambda}_n$ and $\hat{G}_n$ obtained by minimizing (4.19) are strongly consistent. The equivalence of using v_t and $\hat{v}_t$ as $n \to \infty$ is proved in Nicholls and Quinn (1982, Appendix 7.1), where explicit expressions for $\hat{\Lambda}_n$ and $\hat{G}_n$ are also given. ||

For genuine RCA processes with $\Lambda \neq 0$, the conditional prediction error matrix $f_{t|t-1}$ given in (4.17) will be stochastic and condition

D1 of Theorem 3.2 in this case implies that some extra conditions are needed on $\{X_t\}$ to obtain asymptotic normality.

Theorem 4.3: The estimate $\hat{a}_n$ of Theorem 4.2 is asymptotically normal if we assume in addition that $E(X_{ti}^4) < \infty$, i=1,...,d. Similarly $\hat{\Lambda}_n$ and $\hat{G}_n$ are asymptotically normal if in addition to the conditions of Theorem 4.2 we assume $E(X_{ti}^8) < \infty$, i=1,...,d.

Proof: Using (4.17) we have that the matrix R in the condition D1 is given by

$$R = E[F^T(t-1,X)\{F(t-1,X) \Lambda F^T(t-1,X) + G\}F(t-1,X)] \tag{4.21}$$

and it follows from (4.14) and Theorem 3.2 that existence of 4th order moments is sufficient to guarantee asymptotic normality of $\hat{a}_n$.

The matrix U defined in (3.7) is given by $U = E\{F^T(t-1,X)F(t-1,X)\}$ from which the limiting covariance matrix of $n^{\frac{1}{2}}\{\text{vec}(\hat{a}_n) - \text{vec}(a)\}$ can be obtained from (3.8) and (4.21). The covariance matrix is given in a slightly different form in (Nicholls and Quinn 1982, p. 127), but the connection can be established with simple tensor product operations.

When it comes to the parameter vectors $\gamma = \text{vech}(\Lambda)$ and $g = \text{vech}(G)$, Theorem 3.2 can again be used if X_t is replaced by $w_t = \text{vech}(v_t)$ with v_t defined in (4.16). Again it can be shown that the difference between using v_t and $\hat{v}_t$ has no influence on asymptotic results (cf. Nicholls and Quinn 1982, Appendix 7.1).

To ease the comparison with Nicholls and Quinn (1982) we use the notation $Z_t^T = H\{F(t-1,X) \otimes F(t-1,X)\}K^T$. Using (4.20) it is not difficult to show that corresponding to the matrix R of the condition D1 we obtain

$$R^{w} = E\left\{\begin{bmatrix} Z_t \\ I_{d(d-1)/2} \end{bmatrix} f^{w}_{t|t-1} \; [Z_t^T \;\; I_{d(d-1)/2}]\right\} \tag{4.22}$$

where $f^{w}_{t|t-1} = E\{(w_t - \tilde{w}_{t|t-1})(w_t - \tilde{w}_{t|t-1})^T | F^X_{t-1}\}$, and where $\tilde{w}_{t|t-1}$ is given in (4.18). From the definition of w_t and Z_t it is seen that the condition D1 of Theorem 3.2 now requires the existence of 8th order moments of $\{X_t\}$, which is the condition given by Nicholls and Quinn (1982) in their theorems 3.2 and 7.2.

From (4.20) it follows that corresponding to (3.7) we have

$$U^{\hat{w}} = E\begin{bmatrix} Z_t Z_t^T & Z_t \\ Z_t^T & I_{d(d-1)/2} \end{bmatrix}, \tag{4.23}$$

and we can now easily recover the formulae for asymptotic covariance for $(\hat{\gamma}_n, \hat{g}_n)$ given in Nicholls and Quinn (1982, p. 132) from the general formula (3.8).||

This reasoning only gives the asymptotic marginal distributions of $\hat{a}_n$ and $(\hat{\gamma}_n, \hat{g}_n)$, respectively. To prove joint asymptotic normality of $(\hat{a}_n, \hat{\gamma}_n, \hat{g}_n)$ we can consider the penalty function

$$Q_n^2 = \sum_{t=p+1}^{n} |X_t - \tilde{X}_{t|t-1}|^2 + \sum_{t=p+1}^{n} |w_t - \tilde{w}_{t|t-1}|^2 \tag{4.24}$$

and apply basically the same technique (cf. Theorem 2.2).

4.3 Bilinear models

This class of models has received considerable attention recently. We refer to Granger and Andersen (1978), Subba Rao (1981) and Baskara Rao et al (1983) and references therein. A scalar bilinear time series $\{X_t\}$ of type BL(p,q,m,k) is defined for all integers t by the difference equation

$$X_t - a_0 - \sum_{i=1}^{p} a_i X_{t-i} = e_t + \sum_{i=1}^{q} c_i e_{t-i} + \sum_{i=1}^{m} \sum_{j=1}^{k} b_{ij} X_{t-i} e_{t-j} \qquad (4.25)$$

Most of the work in the literature has been concerned with finding stationary F_t^e-measurable solutions to (4.25) and to the evaluation of parameters from the data. We are not aware of a theory of statistical inference for these models, except in rather special cases (cf. Hall and Heyde 1980, Sec. 6.5). Using our general framework we have only been able to treat some special bilinear series, and we will point out the reason for failure in the general case.

Guegan (1983) examines conditions for stationarity for the bilinear model

$$X_t - aX_{t-1} = ce_t + bX_{t-1}e_t + d(e_t^2 - 1) \qquad (4.26)$$

where $\{e_t\}$ is a zero-mean Gaussian white noise with variance 1. Guegan shows that there is an ergodic, strictly and second order stationary solution of (4.26) if $E\{(a+be_t)^2\} < 1$. We can choose m=1 in (3.7) and we have $\tilde{X}_{t|t-1} = aX_{t-1}$ and $f_{t|t-1} = (bX_{t-1}+c)^2 + 2d^2$. It follows directly from Theorems 3.1 and 3.2 that $\hat{a}_n = (\sum_{t=2}^{n} X_t X_{t-1})(\sum_{t=2}^{n} X_{t-1}^2)^{-1}$ is a strongly consistent estimate for a, and that if we assume $E(X_t^4) < \infty$, it is also asymptotically normal. Estimates of b,c, and d can be treated by using Theorems 3.1 and 3.2 on $v_t = \{bX_{t-1}e_t + ce_t + d(e_t^2-1)\}^2$.

The model (4.26) has a very special structure, since no "past" e_t's are allowed. This guarantees that $\{X_t\}$ is a Markov process. We refer to Guegan (1983) for some generalizations.

The difficulties in the general case is illustrated by conditioning on F_{t-1}^X in (4.25). We obtain

$$\tilde{X}_{t|t-1} = a_0 + \sum_{i=1}^{p} a_i X_{t-i} + \sum_{i=1}^{q} c_i E(e_{t-i}|F_{t-1}^X) + \sum_{i=1}^{m}\sum_{j=1}^{k} b_{ij} X_{t-i} E(e_{t-j}|F_{t-1}^X) \quad (4.27)$$

The conditional expectations $E(e_{t-i}|F_{t-1}^X)$ will in general depend nonlinearly on the parameters and on $\{X_s, s \leq t-1\}$, and thus the derivatives with respect to the parameters will in general be infinite expansions in terms of $\{X_s, s \leq t-1\}$ as well. The conditions C1, C3 and D1 essentially require mean square convergence of such expressions and are thus intimately connected with the invertibility problem of bilinear models; i.e. the problem of expressing e_t in terms of a properly convergent nonlinear series of past X_t's. This problem seems very complicated (cf. Granger and Andersen 1978, Ch. 8) and until more progress is made, it appears to be difficult to make substantial headway in conditional least squares estimation of bilinear series using the present framework.

5. A maximum likelihood type penalty function.

In all of the following it will be assumed that the conditional prediction error matrix $f_{t|t-1}$ is nonsingular and that there exists an m such that (3.7) holds.

Corresponding to weighted least squares estimation we introduce the conditional weighted sum of squares penalty function

$$L_n^0 = \sum_{t=m+1}^{n} (X_t - \tilde{X}_{t|t-1})^T f_{t|t-1}^{-1} (X_t - \tilde{X}_{t|t-1}) \triangleq \sum_{t=m+1}^{n} \alpha_t \quad (5.1)$$

We would still like to base our reasoning on the general Theorems 2.1 and 2.2. It was essential in Section 3 that $\{\partial Q_n(\beta^0)/\partial\beta_i, F_n^X\}$ was a zero-mean martingale. We have the following result, where Tr and det are abbreviations for trace and determinant.

Proposition 5.1: Let α_t be as defined in (5.1). Then

$$E\{\frac{\partial\alpha_t}{\partial\beta_i}(\beta^0)|F^X_{t-1}\} = -\mathrm{Tr}\ \{\frac{\partial f_{t|t-1}}{\partial\beta_i}(\beta^0)f_{t|t-1}(\beta^0)\}$$

$$= -\frac{\partial}{\partial\beta_i}\ln\ [\det\ \{f_{t|t-1}(\beta^0)\}] \tag{5.2}$$

Proof: We have

$$\frac{\partial\alpha_t}{\partial\beta_i} = -2(X_t-\tilde{X}_{t|t-1})^T f_{t|t-1}\frac{\partial\tilde{X}_{t|t-1}}{\partial\beta_i} - (X_t-\tilde{X}_{t|t-1})^T f^{-1}_{t|t-1}\frac{\partial f_{t|t-1}}{\partial\beta_i}f^{-1}_{t|t-1}(X_t-\tilde{X}_{t|t-1}) \tag{5.3}$$

Using the definition of $\tilde{X}_{t|t-1}$ and standard rules of the trace operation we have for $\beta = \beta^0$

$$E(\frac{\partial\alpha_t}{\partial\beta_i}|F^X_{t-1}) = -E\{(X_t-\tilde{X}_{t|t-1})^T f^{-1}_{t|t-1}\frac{\partial f_{t|t-1}}{\partial\beta_i}f^{-1}_{t|t-1}(X_t-\tilde{X}_{t|t-1})|F^X_{t-1}\}$$

$$= -\mathrm{Tr}[E\{(X_t-\tilde{X}_{t|t-1})(X_t-\tilde{X}_{t|t-1})^T|F^X_{t-1}\}f^{-1}_{t|t-1}\frac{\partial f_{t|t-1}}{\partial\beta_i}f^{-1}_{t|t-1}]$$

$$= -\mathrm{Tr}\left(\frac{\partial f_{t|t-1}}{\partial\beta_i}\ f^{-1}_{t|t-1}\right) \tag{5.4}$$

The last expression can be written as $-\partial/\partial\beta_i\{\mathrm{Tr}(\ln f_{t|t-1})\}$

However, for a general symmetric matrix A we have $\mathrm{Tr}(\ln A) = \ln(\det A)$ and (5.2) follows. ||

This proposition shows that $\{\alpha_t\}$ is not a martingale difference sequence with respect to $\{F^X_t\}$, but if we introduce an adjustment term corresponding to (5.2) we obtain a penalty function

$$L_n = \sum_{t=m+1}^{n}[\ln\{\det(f_{t|t-1})\} + (X_t-\tilde{X}_{t|t-1})^T f^{-1}_{t|t-1}(X_t-\tilde{X}_{t|t-1})] \triangleq \sum_{t=m+1}^{n}\phi_t \tag{5.5}$$

which has the required martingale property.

If $\{X_t\}$ is a conditional Gaussian process, then L_n coincides with the log likelihood function except for a multiplicative constant. However, in this paper we will not restrict ourselves to Gaussian processes and a likelihood interpretation, but rather view L_n as a

general penalty function which, since it has the martingale property for a general $\{X_t\}$, can be subjected to the kind of analysis described in Sections 2 and 3.

The analysis of L_n will differ in an essential way from that based on conditional least squares only in the cases where $f_{t|t-1}$ is a genuine stochastic process; i.e. when $X_t - \tilde{X}_{t|t-1}$ is not independent of F^X_{t-1}. For the examples treated in detail in Section 4 this is the case only for the RCA processes. More general state space models of this type will be treated in Tjøstheim (1984b). As will be seen, using L_n it is sometimes possible to relax moment conditions on $\{X_t\}$.

5.1 Consistency

We denote by s the number of components of the parameter vector β appearing in $L_n(\beta)$. Due to the presence of $f_{t|t-1}$ in L_n, in general $s > r$ with r defined as in Theorem 3.1.

Theorem 5.1: Assume that $\{X_t\}$ is a d-dimensional strictly stationary and ergodic process with $E(|X_t|^2) < \infty$, and that $\tilde{X}_{t|t-1}(\beta)$ and $f_{t|t-1}(\beta)$ are almost surely three times continuously differentiable in an open set B containing β^0. Moreover, if ϕ_t is defined by (5.5), assume that

E1: $E\left(\left|\frac{\partial\phi_t}{\partial\beta_i}(\beta^0)\right|\right) < \infty$ and $E\left(\left|\frac{\partial^2\phi_t}{\partial\beta_i\partial\beta_j}(\beta^0)\right|\right) < \infty$

for i,j=1,...,s, and where expressions for these derivatives are given in (5.8) and (5.9)

E2: For arbitrary real numbers $a_1,\ldots,a_s$ such that for $\beta=\beta^0$

$$E\left(\left|f^{-\frac{1}{2}}_{t|t-1}\sum_{i=1}^{s} a_i\frac{\partial\tilde{X}_{t|t-1}}{\partial\beta_i}\right|^2\right) + E\left[\left|f^{-\frac{1}{2}}_{t|t-1}\otimes f^{-\frac{1}{2}}_{t|t-1}\sum_{i=1}^{s} a_i\frac{\partial}{\partial\beta_i}\{\mathrm{vec}(f_{t|t-1})\}\right|^2\right] = 0, \quad (5.6)$$

then we have $a_1=a_2=\ldots=a_s=0$,

E3: For $\beta \in B$, there exists a function $H_t^{ijk}(X_1,\ldots,X_t)$ such that

$$\left|\frac{\partial^3\phi_t}{\partial\beta_i\partial\beta_j\partial\beta_k}(\beta)\right| \leq H_t^{ijk} \quad \text{and } E(H_t^{ijk}) < \infty$$

for $i,j,k = 1,\ldots,s$.

Then there exists a sequence of estimators $\{\hat{\beta}_n\}$ minimizing L_n of (5.5) such that the conclusion of Theorem 2.1 holds.

Proof: Due to stationarity and ergodicity and the first part of E1, we have $n^{-1}\partial L_n(\beta^0)/\partial\beta_i \xrightarrow{a.s.} E\{\partial\phi_t(\beta^0)/\partial\beta_i\}$ as $n \to \infty$. However, because of the martingale increment property just demonstrated for $\{\partial\phi_t(\beta^0)/\partial\beta_i\}$ we have $E\{\partial\phi_t(\beta^0)/\partial\beta_i\} = E[E\{\partial\phi_t(\beta^0)/\partial\beta_i|F_{t-1}^X\}] = 0$ and A1 of Theorem 2.1 follows. Similarly, A3 of that theorem follows from E3 and the ergodic theorem.

Using the last part of E1 and the ergodic theorem we have

$$n^{-1}\frac{\partial^2 L_n}{\partial\beta_i\partial\beta_j}(\beta^0) \xrightarrow{a.s.} E\left[E\{\frac{\partial^2\phi_t}{\partial\beta_i\partial\beta_j}(\beta^0)|F_{t-1}^X\}\right] \triangleq V'_{ij} \tag{5.7}$$

It remains to show that E2 implies that the matrix $V' = (V'_{ij})$ is positive definite.

For this purpose we will give explicit expressions for $\partial\phi_t/\partial\beta_i$ and $\partial^2\phi_t/\partial\beta_i\partial\beta_j$, since this will be useful also when checking the conditions E1 and E3. From the definition of ϕ_t in (5.5) and from (5.4) we have

$$\frac{\partial\phi_t}{\partial\beta_i} = \mathrm{Tr}\left(f_{t|t-1}^{-1}\frac{\partial f_{t|t-1}}{\partial\beta_i}\right) - 2\frac{\partial\tilde{X}_{t|t-1}^T}{\partial\beta_i} f_{t|t-1}^{-1}(X_t - \tilde{X}_{t|t-1})$$

$$- (X_t - \tilde{X}_{t|t-1})^T f_{t|t-1}^{-1}\frac{\partial f_{t|t-1}}{\partial\beta_i} f_{t|t-1}^{-1}(X_t - \tilde{X}_{t|t-1}) \tag{5.8}$$

and similarly

$$\frac{\partial^2 \phi_t}{\partial\beta_i \partial\beta_j} = \mathrm{Tr}\left(f^{-1}_{t|t-1} \frac{\partial^2 f_{t|t-1}}{\partial\beta_i \partial\beta_j}\right) - \mathrm{Tr}\left(f^{-1}_{t|t-1} \frac{\partial f_{t|t-1}}{\partial\beta_j} f^{-1}_{t|t-1} \frac{\partial f_{t|t-1}}{\partial\beta_i}\right)$$

$$- 2\frac{\partial^2 \tilde{X}^T_{t|t-1}}{\partial\beta_i \partial\beta_j} f^{-1}_{t|t-1} \; (X_t - \tilde{X}_{t|t-1}) + 2 \frac{\partial \tilde{X}^T_{t|t-1}}{\partial\beta_i} f^{-1}_{t|t-1} \frac{\partial \tilde{X}_{t|t-1}}{\partial\beta_j}$$

$$+ 2\frac{\partial \tilde{X}^T_{t|t-1}}{\partial\beta_i} \; f^{-1}_{t|t-1} \frac{\partial f_{t|t-1}}{\partial\beta_j} f^{-1}_{t|t-1} \; (X_t - \tilde{X}_{t|t-1})$$

$$+ 2 \frac{\partial \tilde{X}^T_{t|t-1}}{\partial\beta_j} \; f^{-1}_{t|t-1} \frac{\partial f_{t|t-1}}{\partial\beta_i} f^{-1}_{t|t-1} (X_t - \tilde{X}_{t|t-1})$$

$$+ (X_t - \tilde{X}_{t|t-1})^T f^{-1}_{t|t-1} \frac{\partial f_{t|t-1}}{\partial\beta_j} f^{-1}_{t|t-1} \frac{\partial f_{t|t-1}}{\partial\beta_i} f^{-1}_{t|t-1} \; (X_t - \tilde{X}_{t|t-1})$$

$$+ (X_t - X_{t|t-1})^T f^{-1}_{t|t-1} \frac{\partial f_{t|t-1}}{\partial\beta_i} f^{-1}_{t|t-1} \frac{\partial f_{t|t-1}}{\partial\beta_j} \; f^{-1}_{t|t-1} \; (X_t - \tilde{X}_{t|t-1})$$

$$- (X_t - \tilde{X}_{t|t-1})^T \; f^{-1}_{t|t-1} \frac{\partial^2 f_{t|t-1}}{\partial\beta_i \partial\beta_j} f^{-1}_{t|t-1} \; (X_t - \tilde{X}_{t|t-1}) \tag{5.9}$$

Since, for a F^X_{t-1}-measurable $d \times d$ matrix function $C(t-1,X)$, we have for $\beta = \beta^0$

$$E\{(X_t - \tilde{X}_{t|t-1})^T f^{-1}_{t|t-1} \; C(t-1,X)(X_t - \tilde{X}_{t|t-1}) | F^X_{t-1}\}$$

$$= \mathrm{Tr}[E\{(X_t - \tilde{X}_{t|t-1})(X_t - \tilde{X}_{t|t-1})^T | F^X_{t-1}\} f^{-1}_{t|t-1} C(t-1,X)]$$

$$= \mathrm{Tr}\{C(t-1,X)\}, \tag{5.10}$$

it is easily verified that

$$E\left(\frac{\partial^2 \phi_t}{\partial\beta_i \partial\beta_j} \; | F^X_{t-1}\right) = \mathrm{Tr}\left(f^{-1}_{t|t-1} \frac{\partial f_{t|t-1}}{\partial\beta_i} f^{-1}_{t|t-1} \frac{\partial f_{t|t-1}}{\partial\beta_j}\right)$$

$$+ 2 \frac{\partial \tilde{X}_{t|t-1}}{\partial\beta_i} f^{-1}_{t|t-1} \frac{\partial \tilde{X}_{t|t-1}}{\partial\beta_j} \tag{5.11}$$

However, using standard rules about tensor products and trace operations we have for $\beta = \beta^0$

$$\mathrm{Tr}(f^{-1}_{t|t-1}\frac{\partial f_{t|t-1}}{\partial \beta_i}f^{-1}_{t|t-1}\frac{\partial f_{t|t-1}}{\partial \beta_j}) = \{\mathrm{vec}(\frac{\partial f_{t|t-1}}{\partial \beta_i}f^{-1}_{t|t-1})\}^T \mathrm{vec}(f^{-1}_{t|t-1}\frac{\partial f_{t|t-1}}{\partial \beta_j})$$

$$= \{\mathrm{vec}(\frac{\partial f_{t|t-1}}{\partial \beta_i})\}^T (f^{-1}_{t|t-1} \otimes f^{-1}_{t|t-1})\mathrm{vec}(\frac{\partial f_{t|t-1}}{\partial \beta_j}) \qquad (5.12)$$

Thus if $a_1,\ldots,a_s$ are arbitrary real numbers, then it follows from (5.11) and (5.12) that

$$\sum_{i=1}^{s}\sum_{j=1}^{s} a_i a_j \; E\{E(\frac{\partial^2 \phi_t}{\partial \beta_i \partial \beta_j} | F^X_{t-1})\} = 2E\left(|f^{-\frac{1}{2}}_{t|t-1}\sum_{i=1}^{s} a_i \frac{\partial \tilde{X}_{t|t-1}}{\partial \beta_i}|^2\right)$$

$$+E\left(|f^{-\frac{1}{2}}_{t|t-1} \otimes f^{-\frac{1}{2}}_{t|t-1}\sum_{i=1}^{s} a_i \mathrm{vec}(\frac{\partial f_{t|t-1}}{\partial \beta_i})|^2\right) \geq 0 \qquad (5.13)$$

Hence the matrix V defined in (5.7) is non-negative definite, and due to the positive definiteness of $f_{t|t-1}$ it now follows from (5.13) and E2 that V is in fact positive definite and the theorem is proved.||

Since some of the components of $\beta = [\beta_1,\ldots,\beta_s]^T$ may be missing from either $\tilde{X}_{t|t-1}$ or $f_{t|t-1}$ (cf. the RCA case in Section 4.2), the condition E2 essentially requires the linear independence of non-zero terms of both $\partial \tilde{X}_{t|t-1}/\partial \beta_i$ and $\partial f_{t|t-1}/\partial \beta_i$. The weighting with $f_{t|t-1}$ is necessary to ensure existence of second moments.

The verification of E1 and E3 must proceed from (5.8) and (5.9), where for $\beta = \beta^0$ the formula (5.10) may be useful. We will see that in the scalar RCA case it is possible to obtain bounds with probability one on quantities like $f^{-1}_{t|t-1}\, \partial f_{t|t-1}/\partial \beta_i$, in which case the bounds in E1 and E3 are not very restrictive.

5.2 Asymptotic normality.

To ease comparison with the results of Section 3 we introduce the matrix U' defined by $U' = \frac{1}{2}V'$, where $V' = (V'_{ij})$ is given by (5.7).

Using (5.11) U' is given for $\beta = \beta^0$ by

$$U' = E\left[\frac{\partial\tilde{X}^T_{t|t-1}}{\partial\beta} f^{-1}_{t|t-1}\frac{\partial\tilde{X}_{t|t-1}}{\partial\beta} + \tfrac{1}{2}\left\{\frac{\partial \mathrm{vec}(f_{t|t-1})}{\partial\beta}\right\}^T f^{-1}_{t|t-1}\otimes f^{-1}_{t|t-1}\frac{\partial \mathrm{vec}(f_{t|t-1})}{\partial\beta}\right] \quad (5.14)$$

Corresponding to Theorem 3.2 we have

<u>Theorem 5.2:</u> Assume that the conditions of Theorem 5.1 are fulfilled and that for $\beta = \beta^0$ and $i,j = 1,\ldots,s$

$$\begin{aligned} F1:\ S_{ij} \overset{\Delta}{=} \tfrac{1}{4}E\Bigg[&\left\{-\mathrm{Tr}\left(f^{-1}_{t|t-1}\frac{\partial f_{t|t-1}}{\partial\beta_i}\right)\mathrm{Tr}\left(f^{-1}_{t|t-1}\frac{\partial f_{t|t-1}}{\partial\beta_j}\right)-\mathrm{Tr}\left(f^{-1}_{t|t-1}\frac{\partial f_{t|t-1}}{\partial\beta_i}f^{-1}_{t|t-1}\frac{\partial f_{t|t-1}}{\partial\beta_j}\right)\right\} \\ &+2\mathrm{Tr}[E\{(X_t-\tilde{X}_{t|t-1})\frac{\partial\tilde{X}_{t|t-1}}{\partial\beta_i}f^{-1}_{t|t-1}(X_t-\tilde{X}_{t|t-1})(X_t-\tilde{X}_{t|t-1})^T|F^X_{t-1}\}f^{-1}_{t|t-1}\frac{\partial f_{t|t-1}}{\partial\beta_j}f^{-1}_{t|t-1}] \\ &+2\mathrm{Tr}[E\{(X_t-\tilde{X}_{t|t-1})\frac{\partial\tilde{X}_{t|t-1}}{\partial\beta_j}f^{-1}_{t|t-1}(X_t-\tilde{X}_{t|t-1})(X_t-\tilde{X}_{t|t-1})^T|F^X_{t-1}\}f^{-1}_{t|t-1}\frac{\partial f_{t|t-1}}{\partial\beta_i}f^{-1}_{t|t-1}] \\ &+\mathrm{Tr}[E\{(X_t-\tilde{X}_{t|t-1})(X_t-\tilde{X}_{t|t-1})^T f^{-1}_{t|t-1}\frac{\partial f_{t|t-1}}{\partial\beta_i}f^{-1}_{t|t-1}(X_t-\tilde{X}_{t|t-1})(X_t-\tilde{X}_{t|t-1})^T|F^X_{t-1}\}\cdot \\ &\cdot f^{-1}_{t|t-1}\frac{\partial f_{t|t-1}}{\partial\beta_j}f^{-1}_{t|t-1}]\Bigg] < \infty \end{aligned} \quad (5.15)$$

Let $S=(S_{ij})$, and let $\{\hat{\beta}_n\}$ be the estimators obtained in Theorem 5.1. Then we have $S_{ij} = \tfrac{1}{4}E\left(\frac{\partial\phi_t}{\partial\beta_i}\frac{\partial\phi_t}{\partial\beta_j}\right) - U_{ij}$ and

$$n^{\frac{1}{2}}(\hat{\beta}_n-\beta^0) \xrightarrow{d} N(0,(U')^{-1} + (U')^{-1}S(U')^{-1}) \quad (5.16)$$

<u>Proof:</u> We use the same technique as in the proof of Theorem 3.2. From the martingale central limit theorem in the strictly stationary ergodic situation and a Cramer-Wold argument, it follows that $n^{-\frac{1}{2}}\partial L_n(\beta^0)/\partial\beta$ has a multivariate normal distribution as its limiting distribution if the limiting covariance of this quantity exists. Using Theorem 2.2 this implies asymptotic normality of $\hat{\beta}_n$ and what remains is to evaluate the covariance matrix.

Since $\{\partial L_n(\beta^0)/\partial\beta_i, F^X_n\}$ is a martingale, it is easy to

verify that

$$n^{-1}E\left\{\frac{\partial L_n}{\partial\beta_i}(\beta^0)\frac{\partial L_n}{\partial\beta_j}(\beta^0)\right\} = n^{-1}\sum_{t=1}^{n} E\left\{\frac{\partial\phi_t}{\partial\beta_i}(\beta^0)\ \frac{\partial\phi_t}{\partial\beta_j}(\beta^0)\right\}$$

$$= E\left[E\left\{\frac{\partial\phi_t}{\partial\beta_i}(\beta^0)\ \frac{\partial\phi_t}{\partial\beta_j}(\beta^0)\,|\,F^X_{t-1}\right\}\right] \qquad (5.17)$$

Using (3.3), (5.8) and (5.10) it is not difficult to show that for $\beta = \beta^0$

$$E\left\{E\left[\frac{\partial\phi_t}{\partial\beta_i}\ \frac{\partial\phi_t}{\partial\beta_j}\,|F^X_{t-1}\right]\right\} = 4(S_{ij} + U'_{ij}) \qquad (5.18)$$

The finiteness of $E\{n^{-\frac{1}{2}}\partial L_n(\beta^0)/\partial\beta_j \cdot n^{-\frac{1}{2}}\partial L_n(\beta^0)/\partial\beta_i\}$ now follows from the assumptions E1 and F1, while the form of the covariance matrix in (5.16) follows from (2.4) and the definition of S and U'. ||

For a conditional Gaussian process we have that X_t is Gaussian conditional on F^X_{t-1} with mean $\tilde{X}_{t|t-1}$ and covariance matrix $f_{t|t-1}$ and for $\beta = \beta^0$ we have

$$E\{(X_{ti}-\tilde{X}_{t|t-1,i})(X_{tj}-\tilde{X}_{t|t-1,j})(X_{tk}-\tilde{X}_{t|t-1,k})\} = 0 \qquad (5.19)$$

for arbitrary components i,j,k=1,...,d. Moreover, from well-known properties of the multivariate normal distribution we have

$$E\{(X_{ti}=\tilde{X}_{t|t-1,i})(X_{tj}-\tilde{X}_{t|t-1,j})(X_{tk}-\tilde{X}_{t|t-1,k})(X_{tm}-\tilde{X}_{t|t-1,m})\}$$

$$= f_{t|t-1,ik}\,f_{t|t-1,jm} + f_{t|t-1,im}f_{t|t-1,jk} + f_{t|t-1,ij}f_{t|t-1,km} \qquad (5.20)$$

for i,j,k,m = 1,...,d. Using this in conjunction with (5.15) it is not difficult to show that S=0 in this case, and (5.16) reduces to

$$n^{\frac{1}{2}}(\hat{\beta}_n-\beta^0) \xrightarrow{d} N(0,(U')^{-1}). \qquad (5.21)$$

In the case where $f_{t|t-1}$ does not depend on the parameter β of interest, we also have S=0 and $U'=E\{\partial\tilde{X}^T_{t|t-1}(\beta^0)/\partial\beta f^{-1}_{t|t-1}\partial\tilde{X}_{t|t-1}(\beta^0)/\partial\beta\}$.

Under the additional assumption of Corollary (3.1) we have

$U' = E[\partial\tilde{X}^T_{t|t-1}(\beta^0)/\partial\beta\{E(f_{t|t-1})\}^{-1}\,\partial\tilde{X}_{t|t-1}(\beta^0)/\partial\beta]$ and estimation using L_n of (5.5) and Q_n of (3.1) essentially gives identical results.

For a scalar process $\{X_t\}$ our general formulae simplifies considerably. We get for $\beta = \beta^0$

$$U'_{ij} = E\left\{\frac{1}{f^2_{t|t-1}}\left\{f_{t|t-1}\frac{\partial\tilde{X}_{t|t-1}}{\partial\beta_i}\frac{\partial\tilde{X}_{t|t-1}}{\partial\beta_j} + \tfrac{1}{2}\frac{\partial f_{t|t-1}}{\partial\beta_i}\frac{\partial f_{t|t-1}}{\partial\beta_j}\right\}\right\} \tag{5.22}$$

and

$$S_{ij} = \tfrac{1}{4}E\left\{\frac{1}{f^4_{t|t-1}}\left[\frac{\partial f_{t|t-1}}{\partial\beta_i}\frac{\partial f_{t|t-1}}{\partial\beta_j}[E\{(X_t - X_{t|t-1})^4|F^X_{t-1}\} - 3f^2_{t|t-1}]\right.\right.$$
$$\left.\left. + 2E\{(X_t - X_{t|t-1})^3|F^X_{t-1}\}\, f_{t|t-1}\left\{\frac{\partial\tilde{X}_{t|t-1}}{\partial\beta_i}\frac{\partial f_{t|t-1}}{\partial\beta_j} + \frac{\partial\tilde{X}_{t|t-1}}{\partial\beta_j}\frac{\partial f_{t|t-1}}{\partial\beta_i}\right\}\right]\right\}, \tag{5.23}$$

where $i,j=1,\ldots,s$. In the next section we will restrict ourselves to the scalar case, since this is the only case considered by Nicholls and Quinn *(1982) for RCA processes.*

6. An example: RCA processes

The method used by Nicholls and Quinn (1982, Ch.4) requires compactness of the region over which the parameter vector is allowed to vary. This necessitates rather restrictive conditions (cf. conditions (ci) - (cii), p. 64 of their monograph). On the other hand the boundedness conditions on the moments are weaker than in the conditional least squares case.

Using our general theoretical framework we are able to dispense with the compactness conditions, while retaining the same weak conditions on the moments. As in Section 4 we assume that conditions are fulfilled so that an ergodic strictly and second order stationary $F^e_t \vee F^b_t$-measurable solution of (4.12), or equivalently (4.13), exists. Moreover, we will again omit the superscript 0 for the true value of

the parameter vector. Finally, it is clear that (3.7) is satisfied with m=p.

In the scalar RCA case we have from (4.14) that

$$F(t-1,X) = [X_{t-1},\ldots,X_{t-p}] \triangleq Y_{t-1}^T \tag{6.1}$$

and, using (4.15), $\tilde{X}_{t|t-1} = Y_{t-1}^T a$, such that $\partial\tilde{X}_{t|t-1}/\partial a_i = X_{t-i}$. Furthermore, from (4.17) it follows that

$$f_{t|t-1} = Y_{t-1}^T \Lambda Y_{t-1} + \sigma^2 , \tag{6.2}$$

where $\sigma^2 = E(e_t^2)$. Corresponding to Theorem 4.2 we have

<u>Theorem 6.1:</u> Let $\{X_t\}$ be a scalar RCA process such that the above stated conditions are satisfied. Assume that $\{e_t\}$ cannot take on only two values almost surely and that Λ is positive definite. Then there exists a sequence of estimators $\{(\hat{a}_n,\hat{\Lambda}_n,\hat{\sigma}_n^2)\}$ minimizing (as described in the conclusion of Theorem 2.1) the penalty function L_n of (5.5) and such that $(\hat{a}_n,\hat{\Lambda}_n,\hat{\sigma}_n^2) \xrightarrow{a.s.} (a,\Lambda,\sigma^2)$.

<u>Proof</u>: We denote by $\lambda_{min} > 0$ the minimum eigenvalue of Λ. It is seen from (6.2) that

$$f_{t|t-1} \geq \lambda_{min} Y_{t-1}^T Y_{t-1} + \sigma^2 \geq \begin{cases} \sigma^2 \\ \lambda_{min} Y_{t-1}^T Y_{t-1} \end{cases} , \tag{6.3}$$

whereas

$$\left|\frac{\partial f_{t|t-1}}{\partial \Lambda_{ij}}\right| = |2X_{t-i}X_{t-j}| \leq Y_{t-1}^T Y_{t-1} \tag{6.4}$$

for i,j=1,...p, and $\partial f_{t|t-1}/\partial\sigma^2 = 1$. It follows from the assumption on $\{e_t\}$ that $\sigma^2 > 0$, and thus $f_{t|t-1}^{-1}$ is well defined and we have from (6.3) and (6.4) that

$$\left| f^{-1}_{t|t-1} \frac{\partial f_{t|t-1}}{\partial \Lambda_{ij}} \right| \leq \frac{2}{\lambda_{min}} \tag{6.5}$$

and

$$\left| f^{-1}_{t|t-1} \frac{\partial f_{t|t-1}}{\partial \sigma^2} \right| \leq \frac{1}{\sigma^2} \tag{6.6}$$

In (5.8) and (5.9) only first order derivatives are non-zero for the RCA case, and it is seen by examining these expressions on a term by term basis that each of the terms involved in evaluating $E(|\partial\phi_t/\partial\beta_i|)$ and $E(|\partial^2\phi_t/\partial\beta_i\partial\beta_j|)$ is bounded by $KE(X_t^2)$ for some constant K. For example for the 6th and 7th term of (5.9) we have with a slight abuse of notation

$$E\left\{\left|2\frac{\partial \tilde{X}_{t|t-1}}{\partial\beta_j} f^{-2}_{t|t-1} \frac{\partial f_{t|t-1}}{\partial\beta_i} (X_t-\tilde{X}_{t|t-1})\right|\right\}+E\left\{\left|(X_t-X_{t|t-1})^2 f^{-3}_{t|t-1} \frac{\partial f_{t|t-1}}{\partial\beta_j} \frac{\partial f_{t|t-1}}{\partial\beta_i}\right|\right\}$$

$$\leq \frac{2}{\sigma^2} \max\left(\frac{2}{\lambda_{min}}, \frac{1}{\sigma^2}\right) E\{|a_j X_{t-j}(X_t-a^T Y_{t-1})|\}$$

$$+ \frac{1}{\sigma^2}\left\{\max\left(\frac{2}{\lambda_{min}}, \frac{1}{\sigma^2}\right)\right\}^2 E\left\{\left|X_t-a^T Y_{t-1}\right|^2\right\}. \tag{6.7}$$

The other terms can be treated similarly and it follows that E1 of Theorem 5.1 is satisfied.

We can use (5.9) to find third order derivatives of ϕ_t. Again, remembering that derivatives of second and higher order for $\tilde{X}_{t|t-1}$ and $f_{t|t-1}$ are zero, we obtain

$$\frac{\partial^3\phi_t}{\partial\beta_i\partial\beta_j\partial\beta_k} = 2f^{-3}_{t|t-1}\frac{\partial f_{t|t-1}}{\partial\beta_i}\frac{\partial f_{t|t-1}}{\partial\beta_j}\frac{\partial f_{t|t-1}}{\partial\beta_k}\{1-3f^{-1}_{t|t-1}(X_t-\tilde{X}_{t|t-1})^2\}$$

$$-2f^{-1}_{t|t-1}\left(\frac{\partial\tilde{X}_{t|t-1}}{\partial\beta_j}\frac{\partial f_{t|t-1}}{\partial\beta_i}\frac{\partial\tilde{X}_{t|t-1}}{\partial\beta_k}+\frac{\partial\tilde{X}_{t|t-1}}{\partial\beta_i}\frac{\partial f_{t|t-1}}{\partial\beta_j}\frac{\partial\tilde{X}_{t|t-1}}{\partial\beta_k}\right.$$

$$\left.+\frac{\partial\tilde{X}_{t|t-1}}{\partial\beta_i}\frac{\partial f_{t|t-1}}{\partial\beta_k}\frac{\partial\tilde{X}_{t|t-1}}{\partial\beta_j}\right)-4f^{-3}_{t|t-1}(X_t-\tilde{X}_{t|t-1})\left(\frac{\partial\tilde{X}_{t|t-1}}{\partial\beta_i}\frac{\partial f_{t|t-1}}{\partial\beta_j}\frac{\partial f_{t|t-1}}{\partial\beta_k}\right.$$

$$\left.+\frac{\partial\tilde{X}_{t|t-1}}{\partial\beta_j}\frac{\partial f_{t|t-1}}{\partial\beta_i}\frac{\partial f_{t|t-1}}{\partial\beta_k}+\frac{\partial\tilde{X}_{t|t-1}}{\partial\beta_k}\frac{\partial f_{t|t-1}}{\partial\beta_i}\frac{\partial f_{t|t-1}}{\partial\beta_j}\right) \tag{6.8}$$

for $i,j,k=1,\ldots,s$. Since $\sigma^2 > 0$ and $\lambda_{min} > 0$, there exists an open set B, that contains the true parameter vector, and is such that the closure of B in the parameter space do not contain $\sigma^2=0$ and $\lambda_{min}=0$. Hence, using exactly the same arguments as above, we find by examining the terms of (6.8) separately that $|\partial^3\phi_t(\beta)/\partial\beta_i\partial\beta_j\partial\beta_k| \leq M|X_t|^2$ for a constant M and where this holds for all $\beta \in B$. Thus, since we assume that $\{X_t\}$ is second order stationary, it follows that condition E3 of Theorem 5.1 is fulfilled.

It remains to verify E2. Let b_i, $i=0,1,\ldots,p$, and b_{ij}, $1\leq i\leq j$ for $j=1,\ldots,p$ and assume that

$$E\{(f^{-\frac{1}{2}}_{t|t-1}\sum_{i=1}^{p} b_iX_{t-i})^2\} + E\{(f^{-1}_{t|t-1}\sum_{i=1}^{j}\sum_{j=1}^{p} b_{ij}X_{t-i}X_{t-j} + b_0)^2\} \tag{6.9}$$

with $f_{t|t-1}$ as in (6.2). Since $f^{-1}_{t|t-1} > 0$, this implies

$$\sum_{i=1}^{p} b_iX_{t-i} \overset{a.s.}{=} 0 \quad\text{and}\quad \sum_{i=1}^{j}\sum_{j=1}^{p} b_{ij}X_{t-i}X_{t-j} + b_0 \overset{a.s.}{=} 0 \tag{6.10}$$

and due to the linear independence properties of RCA processes (cf. proof of Theorem 4.2) it follows that the b's are all zero, and E2 is verified. ||

It is perhaps worth noting that the verification of E1 and E3 given in the above proof holds for any class of processes where the

second and higher order derivatives of $\tilde{X}_{t|t-1}$ and $f_{t|t-1}$ are zero, where $\partial\tilde{X}_{t|t-1}/\partial\beta$ is linear in both β and Y_{t-1}, and where bounds as in (6.5) and (6.6) can be established.

To prove asymptotic normality, according to Theorem 5.2, we have to prove finiteness of S_{ij} with S_{ij} defined as in (5.15). Corresponding to Theorem 4.3 we have

Theorem 6.2: The estimates $(\hat{a}_n, \hat{\Lambda}_n, \hat{\sigma}^2_n)$ obtained in Theorem 6.1 are joint asymptotically normal, if, in addition to the conditions of Theorem 6.1, we assume $E(e_t^4) < \infty$ and $E(b_{ti}^4) < \infty$, $i=1,\ldots,p$.

Proof: We only look at the term

$$E\left[\left| f^{-4}_{t|t-1} \frac{\partial f_{t|t-1}}{\partial\beta_i} \frac{\partial f_{t|t-1}}{\partial\beta_j} E\left\{(X_t-\tilde{X}_{t|t-1})^4 | F^X_{t-1}\right\}\right|\right] \triangleq E[C_{ij}] \tag{6.11}$$

of (5.15). The other terms can be treated likewise.

Using the fact that $\{e_t\}$ and $\{b_t(p)\} = \{[b_{t1},\ldots,b_{tp}]\}$ are independent with $E(e_t) = E\{b_t(p)\} = 0$, we have

$$E\{(X_t-\tilde{X}_{t|t-1})^4|F^X_{t-1}\} = E\{(b_t(p)Y_{t-1} + e_t)^4|F^X_{t-1}\}$$

$$= \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{m=1}^{p} X_{t-i}X_{t-j}X_{t-k}X_{t-m}\, E(b_{ti}b_{tj}b_{tk}b_{tm})$$

$$+ 6\,\sigma^2 \sum_{i=1}^{p}\sum_{j=1}^{p} X_{t-i}X_{t-j}E(b_{ti}b_{tj}) + E(e_t^4) \tag{6.12}$$

From (6.4) and $E(b_{ti}^4) < \infty$, $i=1,\ldots,p$, it follows by successive applications of the Schwarz inequality that

$$\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{m=1}^{p} |X_{t-i}X_{t-j}X_{t-k}X_{t-m}|\, |E(b_{ti}b_{tj}b_{tk}b_{tm})| \leq M_1 p^4 (Y_t^T Y_t)^2 \tag{6.13}$$

and

$$\sum_{i=1}^{p}\sum_{j=1}^{p} |X_{t-i}X_{t-j}|\, |E(b_{ti}b_{tj})| \leq M_2 p^2 Y_t^T Y_t \tag{6.14}$$

for some positive constants M_1 and M_2. Using $E(e_t^4) < \infty$ and (6.3). (6.5) and (6.6) it is seen that C_{ij} defined in (6.11) is bounded with probability one, and thus $E(C_{ij}) < \infty$. The other terms of (5.15) are shown to have a finite expectation using identical arguments, and this completes the proof.||

The asymptotic covariance matrix can now be computed from formula (5.16) and the results of Nicholls and Quinn (1982, Appendix 4.2) can easily be derived. In particular, in the case where $\{b_t(p), e_t\}$ is Gaussian then given F_{t-1}^X, the random variable $X_t - \tilde{X}_{t|t-1} = b_t(p)Y_{t-1} + e_t$ is normal with mean zero and variance $Y_{t-1}^T \Lambda Y_{t-1} + \sigma^2$. It follows that S=0 in (5.16), and the asymptotic covariance matrix is then given by $n^{-1}(U')^{-1}$ with U' as in (5.14). This agrees with the results of Nicholls and Quinn (1982, p. 80).

For the more general processes arising in Kalman type filtering models (Ledolter 1981) the estimation problem is considerably more difficult. These processes are in general nonstationary, and we refer to Tjøstheim (1984b) for a special case.

Acknowledgement

This work was carried out while the author was visiting the Center for Stochastic Processes, Department of Statistics, University of North Carolina. It is a pleasure to acknowledge stimulating discussions with Professor M. Pourahmadi on the problem of existence of stationary solutions.

References

Aase, K.K. (1983): Recursive estimation in non-linear time series models of autoregressive type, J. Roy. Stat. Soc. Ser. B, 45, 228-237.

Anděl, J. (1976): Autoregressive series with random parameters, Math. Operationsforsch. u. Statist., 7, 735-741.

Billingsley, P. (1961): The Lindeberg-Levy theorem for martingales, Proc. Amer. Math. Soc., 12, 788-792.

Bhaskara Rao, M., Subba Rao, T. and Walker, A. M. (1983): On the existence of strictly stationary solutions to bilinear equations, J. Time Series Anal., 4, 95-110.

Granger, C.W.J. and Andersen, A.P. (1978): An Introduction to Bilinear Time Series Models, Vanderhoeck and Ruprecht, Gøttingen.

Guegan, D. (1983): Une condition d'ergodicité pour des modelès bilinéaires à temps discret, C.R. Acad. Sc. Paris, 297, 537-540.

Haggan, V. and Ozaki, T. (1981): Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model, Biometrika, 68, 189-196.

Hall, P. and Heyde, C.C. (1980): Martingale Limit Theory and Its Application, Academic Press, New York.

Harrison, P.J. and Stevens C.F. (1976): Bayesian forecasting (with discussion), J. Roy. Stat. Soc. Ser. B, 38, 205-248.

Jones, D.A. (1978): Non-linear autoregressive processes, Proc. Roy. Soc. London A, 360, 71-95.

Klimko, L.A. and Nelson, P.I. (1978): On conditional least squares estimation for stochastic processes, Ann. Statist., 6, 629-642.

Ledolter, J. (1981): Recursive estimation and forecasting in ARIMA models with time varying coefficients, In Applied Time Series Analysis II, D.F. Findlay Ed., Academic Press, New York, 449-471.

Nicholls, D.F. and Quinn, B.G. (1982): Random Coefficient Autoregressive Models: An Introduction, Lecture Notes in Statistics, 11, Springer, New York.

Ozaki, T. (1980): Non-linear time series models for non-linear random vibrations, J. Appl. Probab., 17, 84-93.

Priestley, M.B. (1980): State-dependent models: A general approach to time series analysis, J. Time Series Anal., 1, 47-71.

Robinson, P.M. (1978): The estimation of a nonlinear moving average model, Stoch. Proc. Appl., 5, 81-90.

Subba Rao, T. (1981): On the theory of bilinear models, J. Roy. Stat. Soc. Ser. B, 43, 244-255.

Tjøstheim, D. (1984a): Recent developments in nonlinear time series modelling, Invited paper, The John B. Thomas Birthday Volume of Papers on Communications, Networks and Signal Processing, Springer, New York, to appear.

Tjøstheim, D. (1984b): Estimation in nonlinear time series models II: Some nonstationary series, Technical Report, Center for Stochastic Processes, Department of Statistics, University of North Carolina, Chapel Hill, N.C. 27514.

Tong, H. and Lim, K.S. (1980): Threshold autoregression, limit cycles and cyclical data (with discussion), J. Roy. Stat. Soc. Ser. B, 42, 245-292.

Tweedie, R.L. (1975): Sufficient conditions for ergodicity and recurrence of Markov chains on a general state space, Stoch. Proc. Appl., 3, 385-403.

END

FILMED

10-84

DTIC